Computer Study of Quantizer Output Spectra

By G. H. ROBERTSON

(Manuscript received January 25, 1969)

This article describes a method for accurately calculating the output spectrum of a quantizer. The method was developed for known expressions defining the output spectrum of an arbitrary quantizer with gaussian input of arbitrary bandshape. Results obtained for a variety of conditions, however, suggest that the calculations are valid even though the input has only a minor gaussian component. When sampling is also used, at the Nyquist rate or a little higher, the quantizing noise folded into the input band is almost flat even when the input bandshape is sharply peaked. When interference at the input is increased, the quantizer (preceded by AGC) appears to operate like an increasingly noisy linear transducer up to a breaking point beyond which its performance (for small signals) degrades rapidly and becomes difficult to analyze.

I. INTRODUCTION

Several authors have described formulas for calculating the output noise spectrum from a quantizer when the input is a gaussian waveform. References 1 through 4 are characteristic and contain representative bibliographies. Evaluation of the resulting expressions is difficult because they contain multiple infinite sums of terms containing Hermite polynomials whose order increases without limit. Consequently simplifying assumptions are made about the input spectrum and quantizer characteristics, or only a few terms are evaluated and the rest assumed negligible, to get results.

This article describes a more fruitful approach in which the Hermite polynomials are evaluated in conjunction with other parts of the expression such that the combination tends to zero as the order increases to infinity. The convergence is slow and many terms are needed to get sufficient accuracy in the noise spectrum. It is possible to get results even when the quantizer is not linear or symmetrical, and for arbitrary input spectrum shapes.

For quantizing steps no greater than σ (the rms gaussian component) an interesting and useful result is that even when the input spectrum is sharply peaked, if the quantized waveform is also sampled uniformly at up to a few times the Nyquist rate for the input band, the resulting quantizing noise appearing within the input band is nearly flat. Many systems can therefore be evaluated quite accurately with much simpler calculations than those needed to define the quantizing noise spectrum.

Study of quantizers having uniform steps less than σ in amplitude show the output spectrum to be practically independent of the location of the gaussian mean if it is at least σ from the overload limit. Consequently, added signals (whose waveform defines the gaussian mean) have a negligible effect on the quantizer output noise as long as tbey do not approach within σ of the limit. A quantizer with many steps activated thus produces a noise spectrum virtually independent of relatively large signals added to the gaussian component.

II. DEFINITION OF QUANTIZER

Figure 1 shows the transfer characteristic of the quantizer where the "staircase" relates the output voltage (ordinate) scale to the input voltage (abscissa) scale. Assuming that the input waveform is gaussian about some arbitrary mean value, the probability that it is Z or more above the mean value is

$$Q_Z = \frac{1}{(2\pi)^{\frac{1}{2}}\sigma} \int_Z^\infty \exp\left(-t^2/2\sigma^2\right) dt. \tag{1}$$

Figure 1 shows that when the input waveform reaches a "riser" of the staircase, the output waveform changes abruptly from the value on one tread to the value of the one on the other side of the riser. For convenience, number the treads and risers starting with 1 at the left. There is one more tread than the number of risers, so if the last riser is k, the last tread is $k + 1$. Let Q_r be the probability that the input waveform is greater than riser r, and the output voltage of step r be W_r. The mean value of the output is

$$S = W_1(1 - Q_1) + W_2(Q_1 - Q_2) + \cdots + W_{k+1}Q_k. \tag{2}$$

The mean squared value is

$$V^2 = W_1^2(1 - Q_1) + W_2^2(Q_1 - Q_2) + \cdots + W_{k+1}^2 Q_k . \tag{3}$$

The variance of the output is

$$V^2 - S^2 = P. \tag{4}$$

Assuming unity impedance, P is the output power after subtracting the component caused by the displacement of the mean value of the input waveform from zero.

III. QUANTIZING NOISE SPECTRUM

Velichkin showed that the correlation function of the quantizer output can be written[4]

$$R_y(\tau) = \sum_{n=1}^{\infty} \left[\sum_{k=1}^{\nu-1} \Delta_k \exp(-a_k^2/2\sigma^2) H_{n-1}\left(\frac{a_k}{\sigma}\right)\right]^2 \frac{R_x^n(\tau)}{2\pi\sigma^{2n} n!}. \tag{5}$$

$R_x(\tau)$ is the input correlation function, σ^2 is the input variance, there are ν treads, Δ_k is the output voltage difference between treads $k+1$ and k, a_k is the input voltage at riser k, and $H_r(z)$ is the Hermite polynomial

$$H_r(z) = (-1)^r \exp(z^2/2) \frac{d^r}{dz^r} [\exp(-z^2/2)]. \tag{6}$$

Also, where $[r/2]$ is the greatest integer $\leq r/2$,[5]

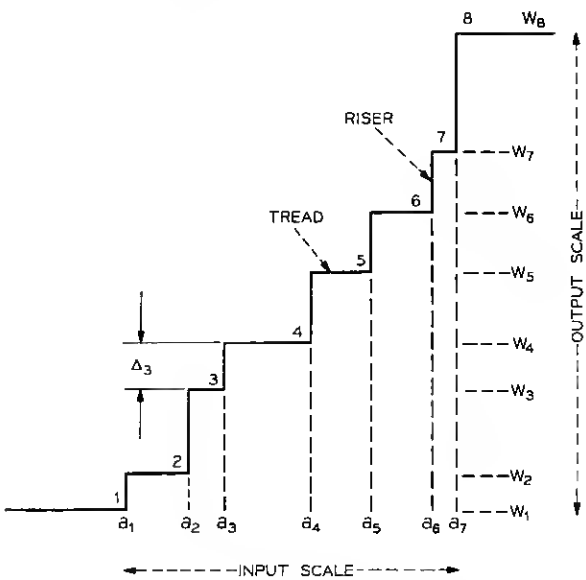

Fig. 1 — Quantizer transfer characteristics.

$$H_r(z) = \sum_{j=0}^{[r/2]} (-1/2)^j z^{r-2j} \frac{r!}{j!\,(r-2j)!}. \tag{7}$$

By the Wiener-Khinchine theorem the power spectrum of the quantizer output is

$$\begin{aligned}\Omega(f) &= 4 \int_0^\infty R_y(\tau) \cos(2\pi f \tau)\, d\tau \\ &= \frac{2}{\pi} \sum_{n=1}^{\infty} \left\{ \sum_{k=1}^{\nu-1} \Delta_k \exp(-a_k^2/2\sigma^2) \frac{H_{n-1}\left(\dfrac{a_k}{\sigma}\right)}{[(n-1)!]^{\frac{1}{2}}} \right\}^2 \frac{1}{n\sigma^{2n}} \\ &\quad \cdot \int_0^\infty R_x^n(\tau) \cos(2\pi f\tau)\, d\tau. \end{aligned} \tag{8}$$

Equation (8) can be written

$$\Omega(f) = \sum_{n=1}^{\infty} \frac{F_n}{\sigma^{2n}} 4 \int_0^\infty R_x^n(\tau) \cos(2\pi f \tau)\, d\tau, \tag{9}$$

in which the quantizing factor terms F_n depend only on the properties of the quantizer and n. When $n = 1$ the component $\Omega_o(f)$ is the input spectrum multiplied by F_1/σ^2. All other n give components whose bandwidth exceeds that of the input [because the integral in equation (9) then represents multiple convolutions of the input band], and their sum $\Omega_e(f)$ may be called the quantizer error spectrum. The quantizer output spectrum is

$$\Omega(f) = \Omega_o(f) + \Omega_e(f). \tag{10}$$

So far only amplitude quantizing has been considered. Sampling, at a rate f_s, is generally also used,* and the output spectrum becomes proportional to[1]

$$\Omega_s(f) = \Omega(f) + \sum_{n=1}^{\infty} \Omega(nf_s \pm f). \tag{11}$$

If f_s is at least twice the highest frequency of the input band, only Ω_e can add more noise by fold-over into the range of the input band.

These results are all known but now follow what are thought to be new contributions enabling equation (9) to be evaluated for an arbitrary choice of input spectrum shape. Equation (9) can be written

---

* The result is independent of which is done first.

$$\begin{aligned}\Omega(f) &= \sum_{n=1}^{\infty} \frac{F_n}{\sigma^{2n}} 2 \int_{-\infty}^{\infty} R_x^n(\tau) e^{-i2\pi f\tau}\, d\tau \\ &= \sum_{n=1}^{\infty} \frac{F_n}{\sigma^{2n}} 2C_{n-1}[s(f)/2], \end{aligned} \tag{12}$$

where $C_{n-1}[s(f)/2]$ is the $(n-1)$th convolution of the cisoid power spectrum $s(f)/2$. The sinusoid power spectrum is $s(f)$, corresponding to the autocorrelation function $R_x(\tau)$.

The significance of equation (12) is that whereas the direct calculation of $R_x^n(\tau)$ may be impossible for an arbitrary spectrum shape, $C_{n-1}[s(f)/2]$ can always be calculated if $s(f)$ is defined. Appendix A describes the methods used to calculate $C_{n-1}[s(f)/2]$ in the computer program written to evaluate equation (9).

If $G(t)$ represents the input waveform, the autocorrelation function at zero lag is

$$\begin{aligned}R_x(0) &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} G(t)^2\, dt \\ &= \sigma^2 + S^2, \end{aligned} \tag{13}$$

where S is the mean value of the input waveform and σ^2 is the variance as used in equation (1). Figure 1 shows that the value of the quantizer output for a given input waveform is independent of the scale on the input axis. For convenience, relabel this scale so that the input mean is zero. Consequently,

$$R_x(0) = \sigma^2. \tag{14}$$

Normalizing the input power that now contains no dc, so that $\sigma^2 = 1$, gives

$$R_x^n(0) = 1 \tag{15}$$

for all n. By the Wiener-Khinchine theorem the total output power P_T is

$$\begin{aligned}P_T &= \int_0^{\infty} \Omega(f)\, df \\ &= R_y(0) \\ &= \frac{1}{2\pi} \sum_{n=1}^{\infty} \frac{1}{n} \left[\sum_{k=1}^{\nu-1} \Delta_k \exp\left(-a_k^2/2\right) \frac{H_{n-1}(a_k)}{[(n-1)!]^{\frac{1}{2}}} \right]^2 \end{aligned} \tag{16}$$

using equation (5). P_T is the same as V^2 given by equation (3) so the

accuracy of computing the quantizing factor terms F_n can be checked by computing the total power by both these methods. The method of equation (3) gives high accuracy very easily, but the F_n terms are needed to compute the quantizer error spectrum $\Omega_e(f)$ of equation (10). Appendix B describes the methods used to compute F_n for values of n up to 10,000, the limit used in the program.

Recall [after equation (9)] that when $n = 1$ the resulting component of $\Omega(f)$ is the input spectrum multiplied by F_1/σ^2, where the gain factor is

$$F_1 = \frac{1}{2\pi}\left[\sum_{k=1}^{r-1} \Delta_k \exp\left(-a_k^2/2\sigma^2\right)\right]^2. \tag{17}$$

When $\sigma^2 = 1$, the total quantizer error power is

$$P_E = P - F_1 , \tag{18}$$

where P is given in equation (4). Both P and F_1 can be computed easily and accurately, so P_E can be determined accurately with little computational effort. Note that this shows P_E to be independent of the input spectrum shape.

A computer program, using the techniques described in Appendixes A and B to compute $\Omega(f)$, simulated the effect of sampling (without holding) by pivoting $\Omega(f)$ about the sampling frequency and its harmonics, and computing the contributions thus folded into the original band. The total P_E is folded into a bandwidth equal to half the sampling frequency; and when the latter was less than a few times the Nyquist rate for the input band, the level of the error component resulting from P_E was nearly flat over the input band even when the input spectrum was sharply peaked.

This result is very useful because the performance of quantizers can now be evaluated quite accurately using only the simple calculations indicated by equations (4) and (17). The error spectrum after sampling was flatter when more levels were used in the quantizer.

IV. SIGNALS ADDED TO INPUT

Signals added to the gaussian noise at the input cause the mean value of the latter to vary according to the signal waveform. Computation shows that under static conditions the gain factor F_1 and the total error power P_E remain nearly constant when the step size is about σ and the mean is no closer than σ to the overload limit. Under these conditions the position of the mean has negligible effect on the shape of the quantizing noise spectrum. Assuming a signal wave-

form uncorrelated with the gaussian noise, and of a magnitude such that the mean rarely approaches within σ of the overload limit, it is thus quite accurate to assume that tbe quantizing error is independent of the signal when the step-to-σ ratio is constant and less tban unity. A sampling rate, up to a few times the high end of the input band, further improves the accuracy of tbis assumption as the quantizing noise then becomes almost flat across the input band even when the input spectrum is sharply peaked.

Assume now that an AGC unit is used to maintain constant power into the quantizer so that the waveform representing the sum of the gaussian component and large signal (interference) very rarely exceeds the overload limits. As the level of the interference increases, the ratio of quantizer step to gaussian rms (rms_g) also increases. Assuming no correlation between the interference and gaussian components, the degradation from quantizing noise can be estimated from the way the parameters F_1 and P_E vary with tbe position of the mean. The greatest variation in these parameters occurs between the values when the mean is at a riser (see Fig. 1) and when it is midway between risers.

Figures 2 and 3 show the results obtained for a 16-level quantizer with a flat input spectrum and with a sharply peaked input spectrum, respectively, in calculations carried out for these limiting cases. Up to a breaking point (where the two curves diverge) tbe quantizer appears to act like a linear but noisy transducer for input signals. Note that the breaking point seems to be independent of the spectrum shape. When the interference level is high enough to cause operation beyond the breaking point, the spectrum becomes difficult to analyze and depends on the interference waveform. At all points on the abscissas of Figs. 2 and 3 below the breaking point, F_1 and P_E were found virtually constant for all positions likely to be occupied by the input mean (determined by the AGC unit). Since the quantizing noise level was flat it was therefore proportional to P_E. The input copy was proportional to F_1; the curves in Figs. 2 and 3 show the ratio of the level of input copy plus quantizing noise to the level of the input copy alone. The degradation these curves indicate, as the interference increases, results from the decreasing ratio of σ to quantizing step size caused by the AGC unit preceding the quantizer.

V. COMPARISON WITH MEASUREMENTS

A sharply peaked spectrum was produced in the laboratory by filtering the output of a noise generator, and the resulting waveform

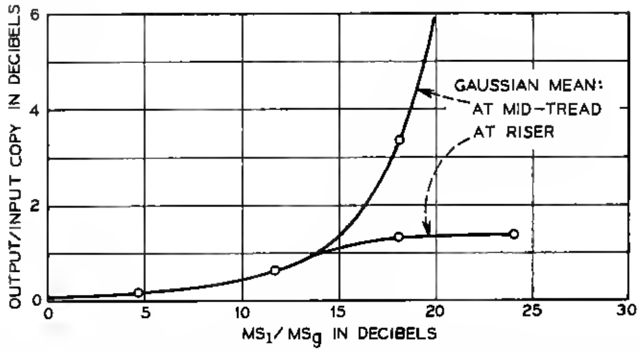

Fig. 2 — Quantizer performance (static) versus relative interference (flat input) for 16-level quantizer with flat input spectrum (gaussian); overload at $3 \times (MS_g + MS_I)^{1/2}$ set by AGC; MS_g = gaussian component (rms)^2; MS_I = interference (rms)^2; output sampled at 3 × high end.

was radically clipped before being submitted to a spectrum analyzer. A 1910-A recording wave analyzer (made by General Radio Company) was used, and several successive traces were superimposed by the recorder as the narrowband (10 Hz) filter was slowly swept across the spectrum. The spectrum before and after clipping were determined in this way; the final results were obtained by drawing a smooth curve through the mean of the superimposed traces. Figure 4, where the solid

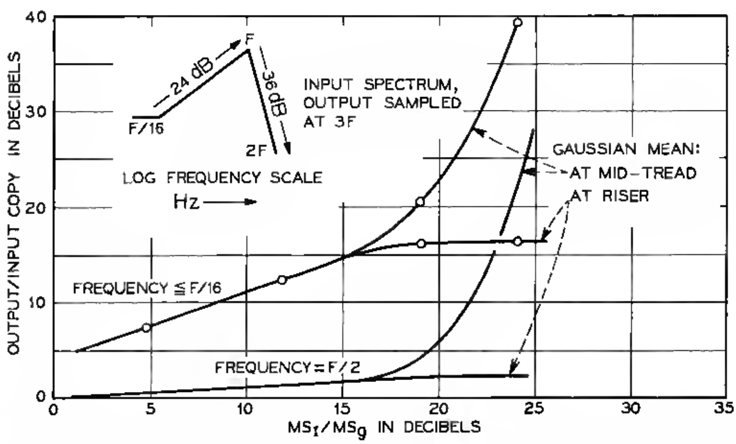

Fig. 3 — Quantizer performance (static) versus relative interference (peaked input) for 16-level quantizer with peaked input spectrum (gaussian); overload at $3 \times (MS_I + MS_g)^{1/2}$ set by AGC.

curve is the computed clipper output spectrum when a copy of the input is given by the dasbed curve, shows the results. Values of the measured output spectrum appear as circles and agree well with the computed curve.

Another check between computed and measured results can be obtained for a uniform step 16-level quantizer. A band of noise, nearly flat from zero to about 330 Hz and falling rapidly at higher frequencies, is added to a sinewave at 160 Hz and passes through an AGC unit before quantization. The quantizer overload limit is set near four times the rms value of the AGC output, and the results are recorded on a magnetic tape for various ratios of the sinewave-to-noise power. In this capacity the sine wave acts as an interfering signal. A computer program processes the tape using a version of the east Fourier transform algorithm to produce estimates of the spectrum level at the quantizer output up to half the sampling rate of 1024 Hz.[6] Since the input spectrum level at 500 Hz is much lower than in the flat part below 300 Hz, the increase in noise level estimated at 500 Hz is taken as a measure of the quantizing noise introduced as the interfering signal increases. Assuming this noise to be flat from 512 Hz to zero it is possible to estimate the degradation in signal-to-noise power suffered by a small signal in the flat part of the input band.

Figure 5 shows the results, as circles superimposed on the solid curves, which are computed for a 16-level quantizer sampled at three times the high end of an input band of noise flat to zero frequency. The quantizer is preceded by an AGC unit and its overload is four times the rms input.

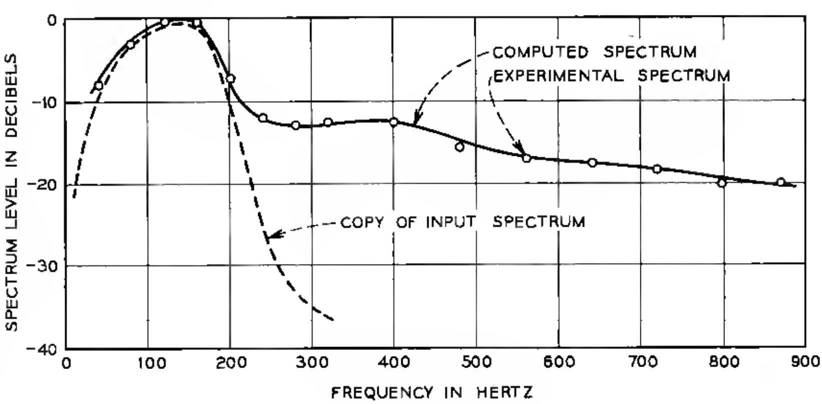

Fig. 4 — Clipper output spectrum.

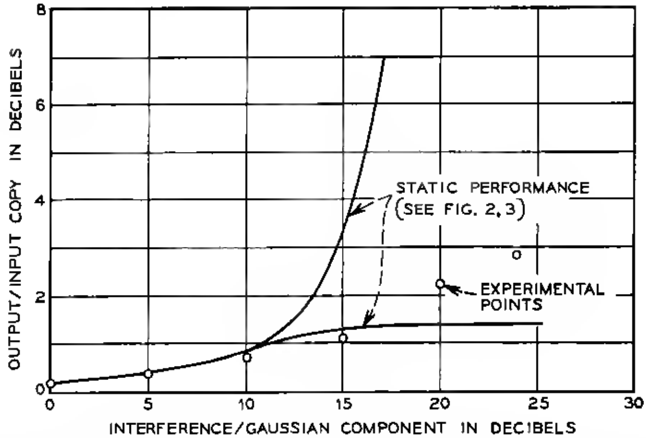

Fig. 5 — Quantizer degradation for flat input spectrum for 16-level quantizer; overload = 4 × input (rms) set by AGC; flat input gaussian component (0 to 330 Hz); sinewave interference (160 Hz); output sampled at 1024 Hz.

VI. CONCLUSION

This article describes a new method of calculating the quantizing noise spectrum when gaussian noise with arbitrary spectrum shape is applied to an arbitrary quantizer. The novelty is not in the form of the expressions that describe the noise spectrum but in the techniques used to compute the results. Applying the method to a sharply peaked spectrum shows that if the output is sampled at the Nyquist rate, or a little higher, the quantizing noise is folded back to cover the input band with almost uniform intensity. A clipper (2-level quantizer) and a 16-level quantizer, preceded by AGC to keep the overload at three times the rms input, operate like noisy but linear transducers for added signals of power less than one tenth and less than twenty times, respectively, that of the broadband background. These useful results indicate that the performance of quantizers under such conditions can be evaluated without the lengthy computations required to delineate the quantizing noise spectrum.

APPENDIX A

Calculating the Input Spectrum Convolutions

The input cisoid spectrum is defined and convolved with itself to calculate $C_{n-1}[s(f)/2]$ when n is small. Because the input spectrum is of finite width, the convolutions tend to take the form of a gaussian

distribution as n increases. Since direct computation of the convolutions becomes very lengthy when n is large, it is profitable to compute a Gram-Charlier approximation instead (see pp. 257–260 of Ref. 5.) This can be done if the moments for the desired convolution can be obtained. The input cisoid spectrum is symmetrical about zero, and is defined up to its limiting bandwidth, so all the moments desired can be computed for it. If the input spectrum shape is normalized so tbat it covers unit area and it is considered to define a probability distribution from which random samples are drawn, the nth convolution is the same as the probability distribution of $(n + 1)$ independent samples of the original distribution.[7] The moments of the nth convolution can thus be obtained from the moments of the input spectrum shape as follows. Since we desire ultimately standardized central moments, note that the standardized central moments for the sum and for the average of N independent samples are the same. Using the appropriate multinomial expansion the general term for the νth such moment is

$$M_\nu = \left(\frac{1}{N^\nu}\right)^{\frac{1}{2}} \sum \left(\frac{\mu_p}{p!}\right)^i \left(\frac{\mu_q}{q!}\right)^j \left(\frac{\mu_r}{r!}\right)^k \frac{\nu!\, N(N-1)\cdots(N-J)}{i!\, j!\, k!}, \tag{19}$$

where

$$\nu = ip + jq + kr, \tag{20}$$

the right side being a partition of ν, and*

$$J = i + j + k - 1.$$

The sum is taken over all the partitions of ν except those containing unity (because the first central moment is zero). The term μ_p is the pth standardized central moment of the original distribution. A program was developed to compute such moments; but since the computation rapidly becomes very lengthy when ν increases, the number of moments used to get the Gram-Charlier approximation was reduced as the convolution order increased. This can be done without undue sacrifice in accuracy since the distribution tends to become gaussian witb increasing convolution order.

APPENDIX B

Calculation of Quantizer Factor Terms F_n

Equation (8) shows that F_n requires computation of terms like

* A partition of ν is a set of positive integers whose sum is ν. The terms i, j, k, p, q, and r are integers.

$$F_{kn} = \exp(-x_k^2/2)H_{n-1}(x_k)/[(n-1)!]^{\frac{1}{2}}, \tag{21}$$

where $H_r(x)$ is a Hermite polynomial for which the recurrence relation exists[8]

$$H_{r+1}(x) = xH_r(x) - rH_{r-1}(x). \tag{22}$$

Therefore

$$F_{k(n+1)} = x_k F_{kn}/n^{\frac{1}{2}} - F_{k(n-1)}[(n-1)/n]^{\frac{1}{2}}. \tag{23}$$

Since $H_0(x) = 1$ and $H_1(x) = x$, from equation (21)

$$F_{k1} = \exp(-x_k^2/2) \tag{24}$$

and

$$F_{k2} = x_k F_{k1}. \tag{25}$$

Therefore, by using equations (23), (24), and (25), a straightforward method exists for finding any F_{kn}. When values are to be calculated using the same x_k and many successive values of n, the programming can be simplified by saving the computed values for n and $(n - 1)$ to be used in equation (23) when the value for $(n + 1)$ is desired. Taking advantage of this way of arranging the computations values were computed for n up to 10,000, enabling determination of the quantizing noise level at greater than 100 times the input bandwidth for a 16-level quantizer. Since recurrence relations like that in equation (23) sometimes result in rapid loss of accuracy, a few values of F_{kn} were computed by an independent method, for high values of n.

Hermite polynomials can be evaluated in terms of confluent hypergeometric functions;[8] a suitable asymptotic formula for these

TABLE I—VALUES OF F_{kn}

x_k	n	Recurrence Relation	Asymptotic Formula
1.0	9999	0.3540125940E-01	0.35401259262E-01
1.0	10,000	0.60060871554397E-01	0.60060871554399E-01
1.0	10,001	−0.34798910623E-01	−0.34798910644E-01
2.0	9999	0.28830572153E-01	0.28830572171E-01
2.0	10,000	0.16057291188981E-01	0.16057291188984E-01
2.0	10,001	−0.28508000965E-01	−0.28508000983E-01
10.0	9,999	−0.62935376617E-12	−0.629353766E-12
10.0	10,000	0.1037121050651E-12	0.1037121050655E-12
10.0	10,001	0.73302922069E-12	0.73302922115E-12

functions was obtained in Ref. 9. Although the asymptotic formula would give adequate accuracy when n is large, the recurrence relation permits much faster evaluations when values are needed over a large range of n. Table I compares a few values of F_{kn} calculated by the recurrence relation and the asymptotic formula. Very good agreement is obtained justifying the use of the recurrence relation.

REFERENCES

1. Bennett, W. R., "Spectra of Quantized Signals," B.S.T.J., *27*, No. 4 (July 1948), pp. 446–472.
2. Bruce J. D., "Correlation Functions of Quantized Signals," *Quarterly Progress Report No. 76,* Massachusetts Institute of Technology Research Laboratory of Electronics, January, 1965, pp. 192–198.
3. Hurd, W. J., "Correlation Function of Quantized Sine Wave Plus Gaussian Noise," IEEE Trans. Inform. Theory, *IT-13,* No. 1 (January 1967), pp. 65–68.
4. Velichkin, A. I., "Correlation Function and Spectral Density of a Quantized Process," *Telecommunications and Radio Engineering,* Part II: Radio Engineering, (July 1962), pp. 70–77.
5. Fry, T. C., *Probability and Its Engineering Uses,* 2nd ed., New York: Van Nostrand, 1965, p. 268.
6. Cooley, J. W., and Tukey, J. W., "An Algorithm for the Machine Calculation of Complex Fourier Series," Math. of Computation, *19,* No. 90 (April 1965), pp. 297–301.
7. Woodward, P. M., *Probability and Information Theory with Applications to Radar,* New York: Pergamon Press, 1960, Section 1.4.
8. Magnus, W., and Oberhettinger, F., *Formulas and Theorems for the Functions of Mathematical Physics,* New York: Chelsea Publishing, 1954, Chapter V, Section 2.
9. Slater, L. J., *Confluent Hypergeometric Functions,* London: Cambridge University Press, 1960, Section 4.5.2.